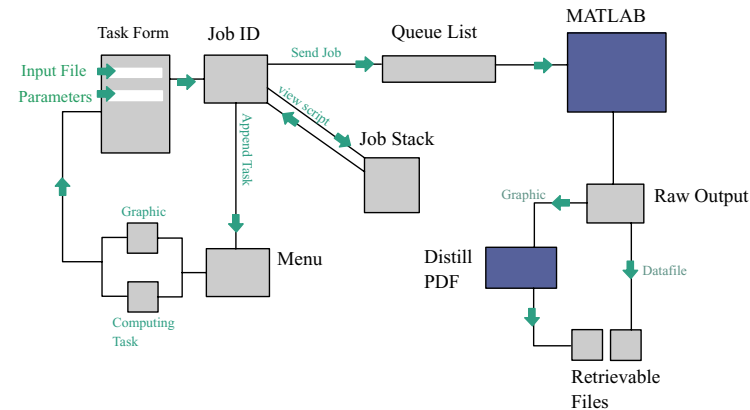


# A New Paradigm for Scientific Computing



David Donoho  
Carrie Grimes  
Stanford University  
April 13, 2001

## **Computing Now:**

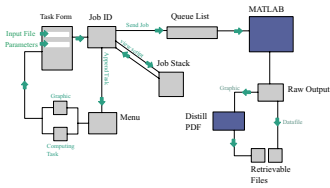
- **single workstation model**
- **techniques developed on local processor**
- **methodology diffusion by portable software**
- **uniform environment for computing**

## **Limitations:**

- **methodology for the single processor model has been extensively mined**
- **frontiers of scientific research pushing capacity of a single processor today**
- **cutting-edge computational experiments restricted to select few with computer power**

## **Task for the Future:**

- **Advance beyond current computing limits while maintaining accessibility to allow new results, algorithms, and implementation to diffuse among other researchers.**



# New Computing Paradigm

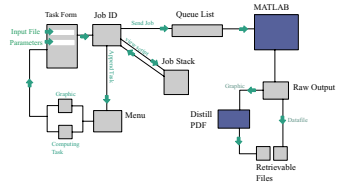
We propose a centralized computing model funded by NSF or Math Institutes with the following purpose:

- **Merge two important facets of research**

- (1) High throughput computing (routines optimized for parallel processors, disk space scavenging, and others)
- (2) Reproducibility and dissemination of research algorithms (shareware, software libraries, interactive web sites)

- **Promote commercial computing trends for science**

- (1) Processor farms for cheap computer power
- (2) Web service providers for information and resource access



# Our Proposal: Accessible Methodology

## ■ Prior to Publication:

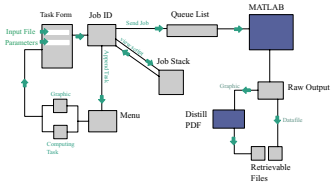
*Researchers use processor farms to do high throughput computing on in-progress research*

## ■ After Publication:

*Algorithms and scripts for published figures available to users for remote computing reproduction.*

## Key Proviso: quid pro quo

To use the system for computing research, you must allow reproducibility afterward



# Web Portal System

## ■ Distributed Computing System

Processor farm (installed at one location) with web interface to access and run research algorithms. Installed basic harmonic analysis functionality

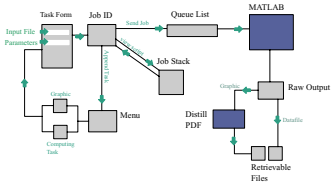
## ■ Three Access Modes

- (1) Reproduce a published figure
- (2) Run an existing algorithm with specified data and parameters
- (3) Upload data and perform basic transforms & functions

## ■ Invariant Web Interface

Stable on all machines, with automatically generated module forms for function interactions. Web Form submission of data and jobs, and web retrieval of results.


# Web-Reproducible Example



↓ Original article, with PDF hot-link or URL reference.

56 Journal of Make-believe Examples

The last twenty years have seen a profound change in the model of scientific publication that has held sway since the founding of the Royal Society in the mid-17<sup>th</sup> century. The rise of computational research in the second half of the 20<sup>th</sup> century has burdened authors with the need both to present results and to make those results (and the algorithms that produced them) accessible to readers and colleagues. With the advent of PC-workstation computing in the 1980's, the central super-computer model of serious computing was phased out in favor of the more accessible local desktop computing. The development of the World Wide Web furthered this decentralized model by allowing not only local research at a workstation, but the wholesale importation of software, algorithms, and results over Web. The result is a model of scientific publication now so common in many scientific disciplines as to qualify as a paradigm: the *methodology-diffusion paradigm*. New complications in the structure of collected data and the



development of more computationally intensive procedures such as 3-D imaging have recently forced the single-workstation to the upper limits of computational capacity. Without easily manageable experimentation for new research techniques, true diffusion of ideas and extension of recently published work is restricted to those with unusual computational capacity. Furthermore, the diversification of computing platforms and software environments used for serious research implies that the transfer of code alone may not be sufficient to enable experimentation even if the computing power is available locally. Unfortunately, the development of processor power in an individual

Fig 2a: Example of estimated student count by major using the priors from section 1. [www.beamlab.org/carrie/example/fig02a](http://www.beamlab.org/carrie/example/fig02a)




Fig 2a: Example of estimated student count by major using the priors from section 1. [www.beamlab.org/carrie/example/fig02a](http://www.beamlab.org/carrie/example/fig02a)

Fig 2a: [View Script Documentation](#)

Fig 2a: [Edit Figure](#)

[Download PDF](#)

↑ Interface generated by RSP. Options include reviewing script documentation, altering stack or parameters, or downloading a PDF of the figure.


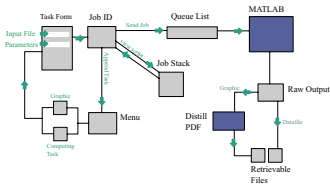


Fig 2a: Example of estimated student count by major using the priors from section 1. [www.beamlab.org/carrie/example/fig02a](http://www.beamlab.org/carrie/example/fig02a)

Functions	Arguments
load_data	edudata .

↑ Web form of the Fig2a computation with editable parameters. Jobs can also be assembled independently. Interface generated by RSP.



# Current Paradigm: Methodology Diffusion Reproducibility as Portable Software

## Wavelab:

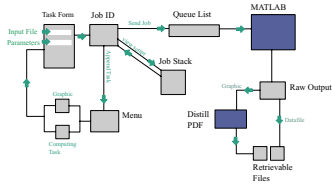
Downloadable code package for Matlab with basic harmonic analysis functionality. Includes documentation, several reproducible book excerpts, and tutorials. Extensions available for Curvelab and Beamlab. [www-stat.stanford.edu](http://www-stat.stanford.edu)  
David Donoho, Stanford Statistics

## SEPLib:

Fortran90 Library + local GUI/Graphics programs for computational image problems in Geophysics. Uses Makerules to construct and re-construct published 'Easily Reproducible Figures.'  
[sepwww.stanford.edu](http://sepwww.stanford.edu)  
Jon Claerbout, Stanford Exploration Project

## Common Characteristics

- (1) target computing platform
- (2) processing/computation run locally
- (3) requires specific commercial software (such as Matlab) or specific compiler environments.



# Why Go Beyond Portable Software?

- **Increasing Data Size and Structural Complexity**

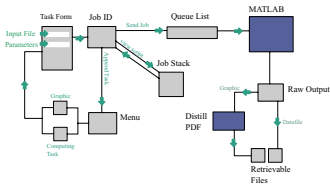
**Examples:** *3D medical imaging, Internet traffic data, Geophysical estimation, Image database problems, Curvelets & Ridgelets on large images*

- **Diversifying Software Base**

**Sample of existing packages:** *Fortran (SEPlib), C (Rainbow), Splus (CART, MART, etc.), Matlab (Wavelab),...*

- **Prevalence of Computationally Costly Methods**

*Cutting-edge estimation methods often imply higher computational cost.*



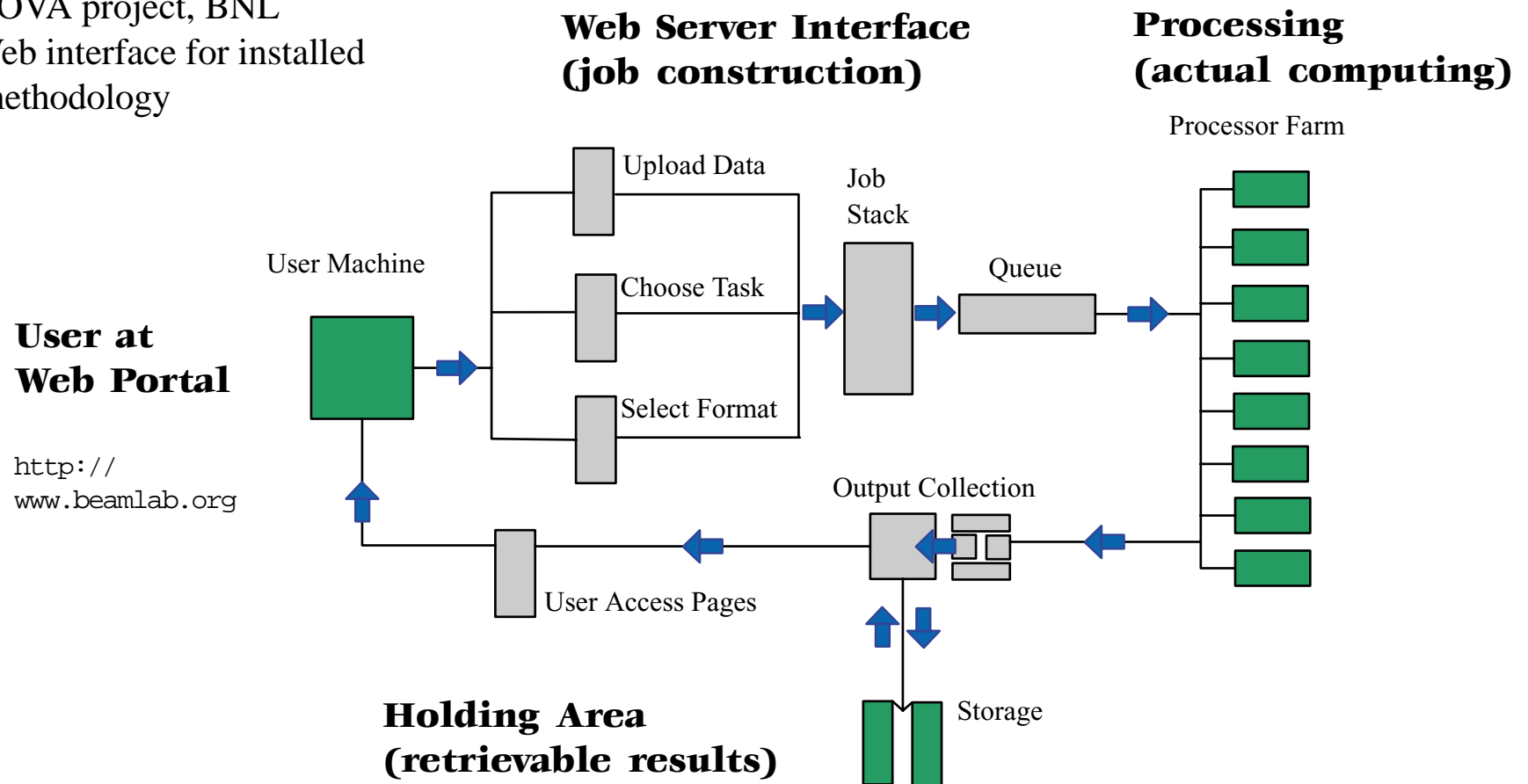
# Advantages of Accessible Methodology

- *availability of serious computing power for research*
- *repository for large or complicated data*
- *standard for ‘reproducibility’*
- *reduces documentation overhead through uniform structure*
- *easy comparison with other techniques*
- *resurrection of old research*

# Proposed Portal System Diagram

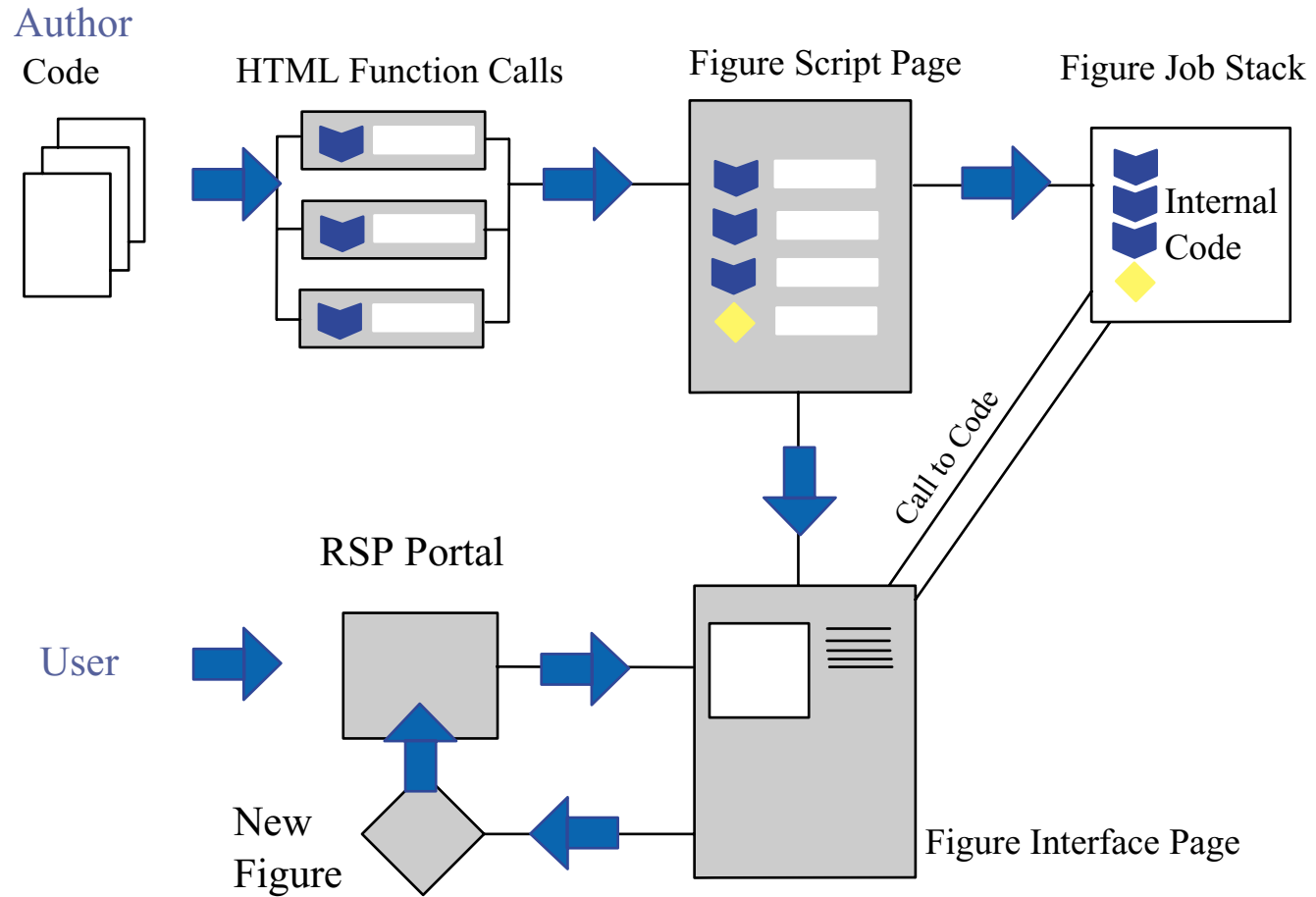
## System Precedents:

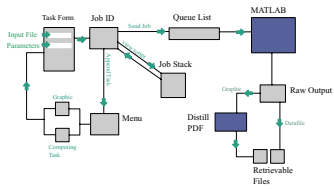
- SETI@home, reverse flow
- NOVA project, BNL  
Web interface for installed methodology



# Innovations

## Interface Updates by Authors





## In the near future:

- Stanford prototype at [www.beamlab.org](http://www.beamlab.org) with a self-contained set of examples and basic harmonic analysis functionality